GESDOR – A Generic Execution Model for Sharing of Computer-Interpretable Clinical Practice Guidelines

Dongwen Wang, PhD[a]; Mor Peleg, PhD[b]; Davis Bu, MD[a]; Michael Cantor, MD[a]; Giora Landesberg, MD, DSc[b]; Eitan Lunenfeld, MD[b]; Samson W. Tu, MS[b]; Gail E. Kaiser, PhD[c]; George Hripcsak, MD, MS[a]; Vimla L. Patel, PhD, DSc[a]; Edward H. Shortliffe, MD, PhD[a]

[a]Department of Medical Informatics, Columbia University, New York, NY 10032
[b]Stanford Medical Informatics, Stanford University, Stanford, CA 94305
[c]Department of Computer Science, Columbia University, New York, NY 10027

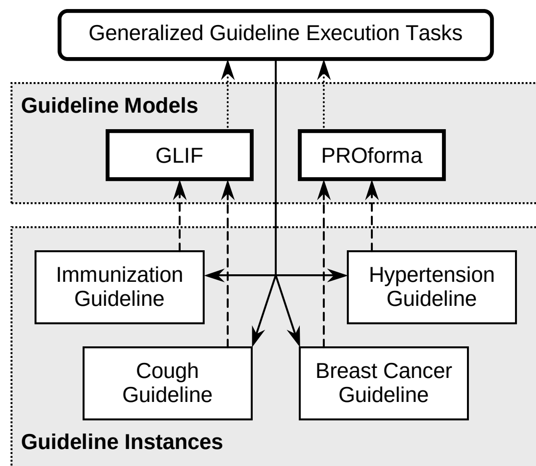
We developed the Guideline Execution by Semantic Decomposition of Representation (GESDOR) model to share guidelines encoded in different formats at the execution level. For this purpose, we extracted a set of generalized guideline execution tasks from the existing guideline representation models. We then created the mappings between specific guideline representation models and the set of the common guideline execution tasks. Finally, we developed a generic task-scheduling model to harmonize the existing approaches to guideline task scheduling. The evaluation has shown that the GESDOR model can be used for the effective execution of guidelines encoded in different formats, and thus realizes guideline sharing at the execution level.

INTRODUCTION

Sharing of computer-interpretable clinical practice guidelines (CPGs) is a critical requirement for guideline development, dissemination and implementation[1]. In addition to conferring cost efficiency in guideline development, guideline sharing leads to improved acceptance of guideline implementation systems, and thus promotes the use of guidelines[2].

One approach to guideline sharing is to develop a universal standard for guideline representation to encode all the guidelines. Considering that no existing guideline representation model is dominant over the others, this approach is impractical at present.

In this study, we propose an alternative approach, the Guideline Execution by Semantic Decomposition of Representation (GESDOR) model, to guideline sharing at the execution level. This approach is based on the observation that the different guideline representation models contain similar execution tasks, which are used to support the implementation of CPGs. According to the GESDOR model, guidelines can be encoded in different formats. A set of generalized guideline execution tasks are extracted from the existing guideline representation models. This set of generalized guideline execution tasks is then used to drive the execution of specific guidelines encoded in different formats. The relationship among the guideline instances, the guideline representation models in which the guideline instances are encoded, and the generalized guideline execution tasks is shown in Figure 1.



mapping between a model and the set of guideline tasks
encoding of a guideline instance in a specific model
execution of a guideline driven by the guideline tasks

Figure 1. The relationship among the guideline instances, the guideline models, and the generalized guideline execution tasks in GESDOR. The guideline instances are encoded in specific representation models, while these models are mapped to the generalized guideline execution tasks. The guideline tasks are then used to drive the execution of the guideline instances encoded in different formats.

METHODS

The GESDOR guideline execution model comprises

(1) a set of guideline representation models, which defines the domain to which the GESDOR guideline execution model can be applied,
(2) a set of generalized guideline execution tasks that are extracted from the existing guideline representation models,
(3) a set of mapping relationships, each of which corresponds to a specific guideline representation

model defined in (1) and provides the semantic links from the elements of that model to the guideline tasks defined in (2), and

(4) a generic task-scheduling model, which harmonizes the existing approaches to task scheduling during guideline execution.

To implement the GESDOR model, the generalized guideline execution tasks need to be extracted first. The mapping relationship between a specific guideline representation model and these guideline tasks needs then to be created. Finally, a generic task-scheduling model needs to be developed to harmonize the existing approaches to task scheduling.

The Generalized Guideline Execution Tasks

To extract the generalized guideline execution tasks, we performed a comprehensive literature review on the existing guideline representation models[3]. Guideline documentation models were used as complements to this review. Two specific guideline models, the 3rd version of the GLIF model (GLIF3)[4] and a variant of the PROforma model (PROforma*)[5], developed and structured as ontologies using the Protégé-2000 knowledge acquisition tool[6], were used as the working templates during this process. Here the PROforma* model inherited most components of the original PROforma model, with the changes only in expression language, cyclic task execution, and patient data definition to simplify the implementation of the GESDOR execution engine. Based on these analyses, we have found a set of generalized guideline execution tasks and a guideline's process structure that are common across different guideline representation models. These generalized guideline execution tasks include (1) the primary tasks, such as *data collection*, *clinical intervention*, *medical decision making*, *patient state verification*, *branching*, *synchronization*, and *subguideline*, which constitute the basic unit of a guideline's process structure, and (2) the auxiliary tasks, such as *criterion evaluation*, *event registration*, and *event invocation*, which are used to support the execution of the primary tasks.

To represent a generalized guideline execution task, we used (1) a set of *input elements*, which define the participants of the task, (2) a set of *output elements*, which define the execution effects of the task, (3) a set of *subtasks*, which define the other guideline execution tasks that are embedded within the task, and (4) a set of *execution constraints*, including *preconditions*, *postconditions*, and *events*, which define the restrictions on the invocation, completion, and triggering of a primary task. The generalized guideline execution tasks were then integrated and organized as an ontology, with each class representing a specific task, a structural element, or an execution constraint, and the slots of the class representing the attributes of that class or its relationships with other classes. We took an incremental approach to the development of this generalized guideline execution task ontology. During this process, we used Protégé-2000 as the knowledge acquisition tool. Detailed description of the development of the generalized guideline execution task ontology can be found elsewhere[7].

Mapping between a Guideline Model and the Generalized Guideline Execution Tasks

The mapping relationship between a guideline representation model and the generalized guideline execution tasks creates the semantic links between them. It is used in the GESDOR model as a set of rules to translate the guideline instances from their original encoding formats to the instances of the guideline tasks that drive the execution of the guidelines.

We assumed that the guideline representation models in this research would be organized as ontologies. As the generalized guideline execution tasks were also arranged as an ontology, the mapping between a guideline model and the generalized guideline execution tasks became the mapping between two ontologies. Accordingly, we defined the mapping relationship at the class layer and the slot layer, with pairs of anchoring classes as the basic units. For this purpose, we developed a mapping model to create the class mapping and the slot mapping between a guideline representation model and the generalized guideline execution tasks. At the class layer, this included the definition of the condition for the class mapping; at the slot layer, this included the specification of the condition for the slot mapping, the use of transformation, and the specific mapping types. Detailed description on this mapping model is beyond the scope of this paper but can be found elsewhere[7].

To facilitate the development and maintenance of the mapping relationship between a guideline representation model and the generalized guideline execution tasks, we developed the GESDOR Ontology Mapping Editor. In addition, we developed a set of guiding principles that assists to make critical decisions when creating the mapping relationships. Based on the mapping model and the guiding principles, using the GESDOR Ontology Mapping Editor as an assisting tool, we had successfully developed the mapping relationship between the GLIF3 guideline representation model and the generalized guideline execution tasks as well as the mapping relationship between the PROforma* guideline representation model and the generalized guideline execution tasks.

The GESDOR Guideline Execution Model

The GESDOR model is built on the approach of guideline execution that was used by GLEE[8], the execution engine for guidelines encoded in the

GLIF3 format. In contrast to GLEE, the GESDOR model uses generalized guideline execution tasks to drive the execution of guidelines. Specifically, guidelines encoded in different formats are stored in a guideline repository, from which they can be retrieved and translated into the instances of the guideline tasks. This translation process is directed by the mapping relationship between the generalized guideline execution tasks and the model in which a guideline is encoded. Once the translation has been completed, the guideline task instances are used by the GESDOR guideline execution engine, along with a generic task-scheduling model that harmonizes the existing approaches to task scheduling, to drive the execution of the guideline. The overall system architecture of the GESDOR model is shown in Figure 2.

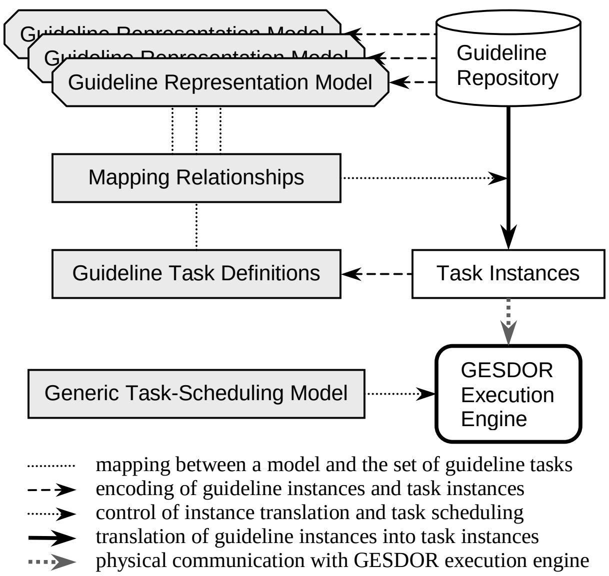


Figure 2. The overall system architecture of the GESDOR guideline execution model.

Study Design

We performed a study to evaluate the effectiveness of the GESDOR guideline execution model, where the model was evaluated as a whole. Specifically, The GLIF3 model and the PROforma* model were selected as the two prototype guideline representation models in the evaluation. The diphtheria-tetanus-pertussis (DTP) series of the childhood immunization guideline published by the Centers for Disease Control and Prevention (CDC)[9] and the cough guideline published by the US Army[10] were selected as the two guideline instances. These two guidelines were then encoded using both the GLIF3 and the PROforma* model. The encoded guideline instances covered all the classes and slots of the two models. To reduce the possible biases in the guideline encoding process, the primary developer of the PROforma* ontology did not directly participate in the encoding of the subject guidelines, and the primary developer of GLEE participated in the encoding of only one of the GLIF3 versions of the two guidelines. For the DTP immunization guideline, 2007 patient cases that had been used previously in a clinical trial on a computerized immunization registry, the EzVac system[11], which implemented the same DTP immunization guideline but not based on any guideline model, were reused in this study. For the cough guideline, domain experts manually created 20 patient cases.

To evaluate the effectiveness of the GESDOR model, we compared the different approaches to the execution of the two subject guidelines. These different approaches included (1) the GESDOR GLIF3 approach, where guidelines were encoded in the GLIF3 format and executed by the GESDOR guideline execution engine, (2) the GESDOR PROforma* approach, where guidelines were encoded in the PROforma* format and executed by the GESDOR guideline execution engine, (3) the GLEE approach, where guidelines were encoded in the GLIF3 format and executed by the GLEE engine, and (4) an ad hoc approach, where guideline encoding and execution were not based on any models. Here the ad hoc approach applied only to the DTP immunization guideline when the EzVac system was used. By comparing the execution results of GESDOR GLIF3 and GLEE, we evaluated the feasibility of the GESDOR model; by comparing the execution results of GESDOR GLIF3 and GESDOR PROforma*, we evaluated the generalizability of the GESDOR model; by comparing the execution results of the ad hoc approach and others, we used the former as an external reference in the evaluation.

The final recommendations generated by the systems and the execution paths that led to the final recommendations were used as the outcomes in the comparison. Here we used three types of execution trace records in the comparison of the execution paths. These types of the trace records included (1) records of the activation of primary tasks, which indicated the temporal sequence that primary tasks were activated, (2) records of the start of primary tasks, which indicated the temporal sequence that primary tasks were actually started, and (3) records of the chaining of primary tasks, which indicated the temporal sequence that primary tasks were chained together (completion of one primary task leading to the activation of other primary tasks during guideline execution). Finally, the clinical validity of the final recommendations was evaluated, using the judgments by physicians as the gold standard.

SUMMARY OF THE RESULTS

When using the four different approaches to execute the DTP immunization guideline, consistent final recommendations were generated in 1978 out of the total 2007 cases (98.56%). In the remaining 29 cases (1.44%), the recommendations generated by GESDOR GLIF3, GESDOR PROforma*, and GLEE were inconsistent with those generated by the EzVac system. The kappa value of 0.98 indicated a high level of agreement of the results.

When using the three different approaches to execute the cough guideline (the ad hoc approach did not apply here), consistent recommendations were generated in all of the 20 cases.

Comparison of the execution paths when GESDOR GLIF3 and GLEE were used to execute the DTP immunization guideline and the cough guideline indicated that the activation traces and the start traces were exactly the same for all the cases of the two guidelines. However, a significant portion of the cases (1946 out of the 2007 cases for the DTP immunization guideline, and all 20 cases for the cough guideline) had inconsistent results when the chaining records were used in the comparison.

Comparison of the execution paths when GESDOR GLIF3 and GESDOR PROforma* were used to execute the DTP immunization guideline and the cough guideline indicated that the activation traces and the start traces were exactly the same for all the cases of the two guidelines. Here the chaining records did not apply, as GLIF3 and PROforma* have different types of primary tasks.

Finally, we used physicians' judgments as the gold standard to evaluate the clinical validity of the final recommendations generated by the systems. For the DTP immunization guideline, all the 29 inconsistent cases and 20 cases that were randomly selected from the 1978 consistent cases were reviewed by two physicians. In the first round of the review, the physician judges did not know the recommendations generated by the systems. Instead, their judgments were based solely on the case descriptions. In this round, the sensitivity and the specificity of GESDOR GLIF3, GESDOR PROforma*, and GLEE (these three systems had the same final recommendations) were 99.71% and 67.65% respectively; and the sensitivity and the specificity of EzVac were 99.43% and 67.48% respectively. To improve the reliability of the judgments, the 5 cases in which the judgments by the physicians were different from any of the four systems were sent back to the physicians for a second review, along with the results generated by the systems this time. In the second round of the review, the sensitivity and the specificity of GESDOR GLIF3, GESDOR PROforma*, and GLEE were 99.80% and 80.74% respectively; and the sensitivity and the specificity of EzVac were 99.53% and 80.55% respectively. Here the EzVac system was used as an external reference to evaluate the other three systems.

For the cough guideline, two physicians reviewed all the 20 cases. The percentage of the correct, acceptable, and wrong diagnoses for case 1 to case 10 were 38.89%, 47.22%, and 13.89% respectively; and the percentage of the correct, acceptable, and wrong diagnoses for case 11 to case 20 were 46.94%, 44.90%, and 8.16% respectively. Here as the first 10 cases were used to tune the encoding of the decision criteria, they were used as a reference to measure the performance of the last 10 cases.

DISCUSSION

The results had shown that the recommendations generated by GESDOR GLIF3, GESDOR PROforma*, and GLEE were exactly the same for all the cases in both guidelines. This means the GESDOR model works well in terms of generating guideline-based recommendations, which are used finally in clinical decision support and thus the most important outcome in the evaluation.

The execution paths of GESDOR GLIF3 and GESDOR PROforma* were exactly the same for all the cases in both guidelines. This means the GESDOR model is generalizable in that it can be applied to different guideline representation models.

The activation traces and the start traces generated by GESDOR GLIF3 and GLEE were consistent for all the cases of both guidelines. Analyses found that the inconsistent results in the comparison of the chaining records were due to the extra information that was added by the generic task-scheduling model of GESDOR. This means that the chaining records should be used (e.g., in implementation of an explanation function associated with a clinical decision support system, where the chaining records play a critical role) conservatively when applying the GESDOR model for guideline execution. It is important to note, however, that this problem of the generic task-scheduling model does not affect the final recommendations generated by the system.

The clinical validity of the final recommendations generated by GESDOR GLIF3 and GESDOR PROforma* reached the level of the reference systems. Specifically, in the execution of the DTP immunization guideline, the sensitivity and the specificity of the systems were at the same level of the EzVac system; in the execution of the cough guideline, the accuracy of the systems when they were applied to the last 10 cases was a little better than that when they were applied to the first 10 cases.

Process modeling tools had been used previously to implement care plans[12]. The GESDOR model is different from previous approaches in that it focused on the process-centered knowledge management, with the generalized guideline execution task ontology as a process-oriented reorganization of the guideline execution knowledge that are common across different guideline models.

Several ontology mapping models and tools had been developed previously for different purposes[13,14]. The ontology mapping model in GESDOR is different from previous approaches in that (1) it focuses on the instance translation directed by model-level mapping, and (2) it has its own languages for specification of slot mapping and mapping condition, and thus provides flexibility in the development of the mapping model to facilitate ontology mapping.

The GESDOR model provides connections among different guideline representation models, similar to the function of the UMLS to bridge different controlled medical terminologies[15]. As a long-term goal, with more and more guideline representation models included into the application domain of GESDOR, a comprehensive standard of guideline representation will be able to be developed and widely accepted.

In this study, we assume that PROforma* uses the same expression language as that in GLIF3. For models using different expression languages, we believe that the general principle of GESDOR still applies, although its effectiveness needs to be evaluated further in those cases. Ideally, the GESDOR model should be tested with guidelines encoded in their original formats. We plan to request additional resources to further investigate the feasibility and the generalizability of the GESDOR model.

CONCLUSION

The GESDOR model can be used for the effective execution of guidelines that are encoded in different formats, and thus realizes guideline sharing at the execution level. GESDOR's chaining records should be used conservatively.

Acknowledgement

This work is supported by grant LM06594 to the InterMed Collaboratory from the National Library of Medicine, the US Army, and the Agency for Healthcare Research and Quality. We gratefully acknowledge the contributions by Bob Greenes, Aziz Boxwala, and others in the InterMed Collaboratory, and the helpful comments from John Fox.

References

1. Shortliffe EH, Patel VL, Cimino JJ, Barnett GO, Greenes RA. A study of collaboration among medical informatics research laboratories. Artif Intell Med. 1998;12(2):98-123.
2. Zielstorff RD. Online practice guidelines: issues, obstacles, and future prospects. J Am Med Inform Assoc. 1998;5(3):227-36.
3. Wang D, Peleg M, Tu SW, Boxwala AA, Greenes RA, Patel VL, Shortliffe EH. Representation primitives, process models and patient data in computer-interpretable clinical practice guideline: a literature review of guideline representation models. Int J Med Inform. 2002;68(1-3):59-70.
4. Peleg M, Boxwala AA, Ogunyemi O, Zeng Q, Tu S, Lacson R, Bernstam E, Ash N, Mork P, Ohno-Machado L, Shortliffe EH, Greenes RA. GLIF3: the evolution of a guideline representation format. Proc AMIA Annu Symp. 2000:645-9.
5. Fox J, Johns N, Rahmanzadeh A. Disseminating medical knowledge: the PROforma approach. Artif Intell Med. 1998;14(1-2):157-81.
6. Musen MA, Fergerson RW, Grosso WE, Noy NF, Crubezy M, Gennari JH. Component-based support for building knowledge-acquisition systems. Conference on Intelligent Information Processing (IIP 2000) of the International Federation for Information Processing World Computer Congress (WCC 2000). 2000.
7. Wang D. A generic execution model for sharing of computer-interpretable clinical practice guidelines [dissertation]. New York (NY): Columbia Univ.; 2003.
8. Wang D, Shortliffe EH. GLEE – a model-driven execution system for computer-based implementation of clinical practice guidelines. Proc AMIA Symp. 2002:855-9.
9. Centers for Disease Control and Prevention. General recommendations on immunization recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep. 1994;43(RR01):1-38.
10. Whitlock W. "Cough", Army Cough Guideline: Special Operations Forces Medical Handbook. Teton NewMedia, 2001.
11. Wang D, Jenders RA, Dasgupta B. Translating national childhood immunization guidelines to a computer-based reminder recall system within an immunization registry. Proc AMIA Annu Symp. 1999:994-8.
12. Lee W, Kaiser GE, Clayton PD, Sherman EH. OzCare: a workflow automation system for care plans. Proc AMIA Annu Symp. 1996:577-81.
13. Noy NF. Musen MA. Anchor-PROMPT: using non-local context for semantic matching. Workshop on Ontologies and Information Sharing at the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001), Seattle, WA. 2001.
14. Crubezy M. Ontologies and problem-solving methods. In: Handbook on Ontologies in Information Systems. Springer. 2003.
15. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993;32(4):281-291.